# Cultural variability in the effects of question design features on respondent comprehension

Johnson, Timothy P.; Cho, Young Ik; Holbrook, Allyson; O'Rourke, Diane; Warnecke, Richard; Chávez, Noel

# CULTURAL VARIABILITY IN THE EFFECTS OF QUESTION DESIGN FEATURES ON RESPONDENT COMPREHENSION

*TIMOTHY P. JOHNSON*, YOUNG IK CHO, ALLYSON HOLBROOK, DIANE O'ROURKE, RICHARD WARNECKE & NOEL CHÁVEZ*

To identify the characteristics of common health survey questions that may be associated with cross-cultural variability in question comprehension, health survey interviews with respondents representing four distinct cultural subgroups in the United States (non-Hispanic White, African American, Mexican American and Puerto Rican) were analyzed via behavior coding. Using survey responses as the unit of analysis (n=13,514), nested within survey respondents (n=345) and survey questions (n=42), hierarchical linear modeling (HLM) was employed to examine the effects of four questionnaire design features on cultural variations in question comprehension difficulties. Question response format, question length, question reading level and level of abstraction were each found to have main effects on respondent comprehension. Respondent culture was found to moderate the effects of response format, question length and reading levels. Several question design strategies that reduce overall comprehension difficulty also increase cross-cultural disparities in this regard.

## 1 Introduction

Respondent culture is now generally understood to influence the comprehension and interpretation of many health survey questionnaires (D'Andrade et al., 1972; Jenkins, 1988; Johnson et al., 1997; Meredith & Siu, 1995; Morse & Morse, 1988; Teresi et al., 2001; Warnecke et al., 1997). Culture-based variations in question comprehension may contribute to differential response artifacts that are erroneously interpreted as cultural differences or disparities in health beliefs, behaviors, and/or conditions (Johnson et al., 1996). Several

* This study was supported in part by Cooperative Agreement #U83/CCU508663 from the National Center for Health Statistics, U.S. Centers for Disease Control and Prevention.

approaches have been taken to address this problem in practice. A variety of new questionnaire translation strategies, for example, have been proposed in recent years to address cross-group differences in question comprehension (Harkness, 2003; Sperber, Devellis & Boehlecke, 1994). Several sets of question wording principles have also been proposed as guidelines for improving comprehension equivalence across cultural groups (Brislin, 1986; McKay et al., 1996; Smith, 2004). Little research, however, is currently available that can provide guidance to researchers regarding other survey question design features that might be useful in reducing comprehension variability when conducting health surveys in culturally heterogeneous environments, such as the United States.

Ironically, numerous question design features are known to be associated with respondent comprehension (Groves et al., 2004; Tourangeau, Rips & Rasinski, 2000). Several of these are believed to reflect question difficulty and/or complexity, including question length, reading level, abstraction level, and response format (Andrews, 1984; Bradburn & Miles, 1979; Bradburn & Sudman, 1979; Just & Carpenter, 1992; Knäuper et al., 1997; Laurent, 1972). The cross-cultural utility of these various question elements, however, have yet to be explored.

The purpose of this study is to investigate each of these common dimensions of health survey questions in order to assess any cultural variability in respondent comprehension that may be associated with each. To accomplish this, behavioral coding (Fowler, 1995) is applied to measure comprehension difficulties across 345 health survey interviews with a culturally diverse sample of U.S. respondents and a range of 42 survey questions. These survey data and behavioral codes offer to provide an assessment of the degree to which the effects of question design features do or do not operate in similar manners across race/ethnic groups.

## 2 Methods

With respondent consent, a total of 345 in-person laboratory interviews were tape-recorded. Using race/ethnicity as a proxy measure of respondent culture, four groups of respondents were examined: African Americans (n=86), Mexican Americans (n=101), Puerto Ricans (n=74), and non-Hispanic Whites (N=84) residing in the Chicago Metropolitan Area. Respondents were recruited via advertisements in local media and ranged in age from 18-53. The interviews averaged approximately an hour in length and all were conducted in English.

### 2.1 Survey questions

The survey instrument included 42 substantive health-related questions selected from national health surveys conducted in the United States. Among the surveys from which items were selected were the National Health Interview Survey (NHIS), the Behavioral Risk Factor and Surveillance Survey (BRFSS) and the National Household Survey of Drug Abuse (NHSDA). Items were selected to represent a variety of topics, question types and formats. The specific wording of all questions is available from the authors. Following each interview, respondents completed a brief inventory of demographic questions.

The 42 survey questions were classified along four dimensions: question length, reading difficulty level, response format, and abstraction level. The length of each question was measured by total number of words. The reading level of each question was graded using Flesch-Kincaid scores (Flesch, 1979). Three response formats were included: those asking for numeric values (e.g., number of times exercise, age first drank alcohol) (n = 13); those for which the respondent could answer "yes" or "no" (n = 9); and those employing vague quantifiers as response categories (e.g., "excellent-good-fair-poor," "strongly agree-agree-disagree-strongly disagree") (n = 20).

Three levels of abstraction were identified. Two of the authors independently classified each question as "most abstract," "somewhat abstract," or "least abstract." Results were subsequently compared and differences discussed and reconciled. Abstract items were defined á priori as those for which the major concept introduced by the question was not grounded in physical reality (n = 11 of the questions examined). Those items classified as "least abstract" were those for which the major concept introduced in the question was grounded in physical reality (n = 17 questions). The remaining 14 items were classified as "somewhat abstract."

### 2.2 Response coding

Audio-tapes were reviewed and respondent reactions to each of the 42 substantive survey questions were coded using a behavioral coding scheme previously reported (Oksenberg, Cannell & Kalton, 1991) and modified for this study. A graduate assistant who was trained and supervised by one of the authors coded a total of 13,514 respondent answers. A random sample of 24 tapes was coded by both persons, revealing an inter-rater agreement of 98.1 percent. Table 1 defines five specific respondent behavior codes that were classified as indicative of comprehension difficulty. An exploratory factor analysis was conducted to evaluate the dimensionality of these five behavior codes. All five behaviors loaded strongly on a single factor. Survey questions that elicited one or more of these behavior codes were consequently defined as producing comprehension difficulty for the respondent in question. Overall, comprehension difficulties were associated with 9.4 percent of the 13,514 survey responses analyzed.

**Table 1 Respondent Behavior Codes Used to Represent Comprehension Difficulties**

1. *Clarification (Unspecified):* Respondent indicates uncertainty about question, but it is unclear as to whether the problem is related to the construct or the context.
2. *Clarification (Construct):* Respondent asks for repeat or clarification of question, or makes a statement indicating uncertainty about question ***meaning*** (i.e., "what do you mean by depressed?").
3. *Clarification (Context):* Respondent indicates s/he understands the meaning of the construct, but indicates uncertainty about question meaning within the context of the question as stated (i.e., "what do you want to know about being depressed?").
4. *Clarification (time frame):* Respondent indicates uncertainty about the question's time frame.
5. *Clarification (rewording):* Respondent rephrases the question before answering.

Descriptive information for each question level and respondent level variable is presented in Table 2.

**Table 2 Question and Respondent Variables**

| Variables | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| *Question Characteristics* | | Level 1 *(N=13,514)* | | |
| Comprehension Difficulty | .09 | .29 | 0 | 1 |
| Abstraction Level | | | | |
| Abstract | .27 | .45 | 0 | 1 |
| In-between | .35 | .48 | 0 | 1 |
| Concrete | .38 | .49 | 0 | 1 |
| Question Length | 21.80 | 10.52 | 5 | 46 |
| Readability Level | 6.84 | 2.81 | 1.50 | 12.0 |
| Response Type | | | | |
| Numeric | .28 | .45 | 0 | 1 |
| Yes/No | .22 | .41 | 0 | 1 |
| Vague quantifiers | .50 | .50 | 0 | 1 |
| *Respondent Characteristics* | | Level 2 *(N=345)* | | |
| Education | 3.58 | 1.16 | 1 | 6 |
| Age | 32.08 | 8.28 | 18 | 53 |
| Gender (Male) | .51 | .50 | 0 | 1 |
| Race/Ethnicity | | | | |
| White | .24 | .43 | 0 | 1 |
| African American | .25 | .43 | 0 | 1 |
| Mexican American | .29 | .46 | 0 | 1 |
| Puerto Rican American | .21 | .41 | 0 | 1 |

### 2.3 Analysis

In order to estimate two-level hierarchical linear models, HLM6 (Raudenbush et al., 2004) was employed. A multi-level modeling strategy enabled an accounting of variance attributed to individual differences (i.e., responses are nested within subjects) and factors associated with individuals as well as questionnaire characteristics. As depicted below, a general conditional model is composed of two sets of equations: equation 1 at the response level, and equation 2 at the respondent level. Since the outcome variable (comprehension difficulty) is binary (problem=1; no problem=0), the model predicts the expected log-odds of the outcome at the first level using a logit link function (Hedeker & Gibbons, 1994). At the first level, expected log-odds of the comprehension difficulty [Log [P/(1-P)]] are predicted by questionnaire characteristics, including abstraction level, question length, reading level and response format. At level-2, intercept at level *1*($B_0$) is modeled as a function of level-2 predictors (respondent demographics) controlling for the random variability ($\mu_0$) across individual respondents.

**Equation 1:** Level-1 Model:

$$\text{Log [P/(1-P)]} = \beta_0 + \beta_1(\text{Most Abstract}) + \beta_2\,(\text{Somewhat Abstract}) + \beta_3\,(\text{Question Length}) + \beta_4\,(\text{Reading Level}) + \beta_5\,(\text{Yes-No Responses}) + \beta_6\,(\text{Vague quantifier Responses})$$

**Equation 2:** Level-2 Model:

$$\beta_0 = \gamma_{00} + \gamma_{01}\,(\text{Education}) + \gamma_{02}\,(\text{Age}) + \gamma_{03}\,(\text{Male}) + \gamma_{04}\,(\text{African American}) + \gamma_{05}\,(\text{Mexican American}) + \gamma_{06}\,(\text{Puerto Rican}) + \mu_0$$

In addition, the effects of level-1 factors are predicted by race/ethnicity to explore interaction effects between level-1 question characteristics and the respondent's race/ethnicity. These effects are estimated in a second model that employs equation 1 and equation 3 as follows:

**Equation 3:** Level-2 Model to examine interaction effects:

$$\beta_0 = \gamma_{00} + \gamma_{01}\,(\text{Education}) + \gamma_{02}\,(\text{Age}) + \gamma_{03}\,(\text{Male}) + \gamma_{04}\,(\text{African American}) + \gamma_{05}\,(\text{Mexican American}) + \gamma_{06}\,(\text{Puerto Rican}) + \mu_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11}\,(\text{African American}) + \gamma_{12}\,(\text{Mexican American}) + \gamma_{13}\,(\text{Puerto Rican American})$$

$$\beta_2 = \gamma_{20} + \gamma_{21}\,(\text{African American}) + \gamma_{22}\,(\text{Mexican American}) + \gamma_{23}\,(\text{Puerto Rican American})$$

$$\beta_3 = \gamma_{30} + \gamma_{31}\,(\text{African American}) + \gamma_{32}\,(\text{Mexican American}) + \gamma_{33}\,(\text{Puerto Rican American})$$

$$\beta_4 = \gamma_{40} + \gamma_{41}\,(\text{African American}) + \gamma_{42}\,(\text{Mexican American}) + \gamma_{43}\,(\text{Puerto Rican American})$$

$$\beta_5 = \gamma_{50} + \gamma_{51}\,(\text{African American}) + \gamma_{52}\,(\text{Mexican American}) + \gamma_{53}\,(\text{Puerto Rican American})$$

## 3 Results

Table 3 presents HLM model results for the main effects of person level and question level characteristics on question comprehension, and Table 4 presents interaction effects between race/ethnicity and each question level indicators. Among the respondent characteristics examined, only race/ethnicity was found to be independently associated with question comprehension problems. Members of minority groups (i.e., African-American, Mexican-American and Puerto Rican respondents) were each more likely to express comprehension difficulties when compared with non-Hispanic white respondents. Respondent age, gender, and education were not related to difficulties.

**Table 3 HLM Estimates of Main Effects of Individual and Question-Level Characteristics on Comprehension Difficulty**

| | Coefficient | (SE) |
|---|---|---|
| *Effects of the individual characteristics* | | |
| Intercept | -2.98*** | (0.26) |
| Education | -0.05 | (0.04) |
| Age | 0.01 | (0.01) |
| Gender (Male) | 0.05 | (0.09) |
| Race/Ethnicity (Ref=White) | | |
| African American | 0.30* | (0.12) |
| Mexican American | 0.50*** | (0.12) |
| Puerto Rican American | 0.40** | (0.14) |
| *Effects of questionnaire characteristics* | | |
| Abstraction Level (Ref= Least Abstract) | | |
| Most Abstract | 0.73*** | (0.09) |
| Somewhat Abstract | 0.12 | (0.08) |
| Question Length | 0.01*** | (0.00) |
| Reading Difficulty Level | 0.07*** | (0.01) |
| Response Format (Ref= Numeric) | | |
| Yes/No | -1.28*** | (0.10) |
| Vague Quantifier | -1.20*** | (0.08) |

*p<.05; **p<.01; *** p<.001.

Each of the four question characteristics was found to be independently associated with respondent comprehension. The effects of three were found to vary across race/ethnic groups. There was a main effect of level of question abstraction. Not surprisingly, those items classified as “most abstract” produced more comprehension difficulties, compared to items designated as “least” abstract. No differences were found between items classified as “moderately” and “least” abstract. The effects of question abstraction did not vary by race/ethnicity as shown in Table 4.

**Table 4 HLM Estimates of Interaction Effects of Race/Ethnicity and Question Characteristics on Respondent Comprehension Difficulty**

| | Coefficient | (SE) |
|---|---|---|
| *Effects of individual characteristics* | | |
| Intercept | -3.37*** | (0.30) |
| Education | -0.05 | (0.04) |
| Age | 0.01 | (0.01) |
| Gender (Male) | 0.05 | (0.09) |
| Race/Ethnicity (Ref=White) | | |
| African American | 0.65** | (0.24) |
| Mexican American | 1.05*** | (0.23) |
| Puerto Rican American | 0.83** | (0.25) |
| *Interaction effects between questionnaire characteristics and respondent race/ethnicity* | | |
| Abstraction Level (Ref=Somewhat Abstract) | | |
| Most Abstract | | |
| Intercept (White) | 0.81*** | (0.20) |
| African American | -0.48 | (0.26) |
| Mexican American | 0.08 | (0.26) |
| Puerto Rican American | 0.11 | (0.27) |
| Moderately Abstract | | |
| Intercept (White) | 0.17 | (0.16) |
| African American | -0.19 | (0.23) |
| Mexican American | 0.05 | (0.22) |
| Puerto Rican American | -0.05 | (0.22) |
| Question Length | | |
| Intercept (White) | 0.02*** | (0.01) |
| African American | -0.00 | (0.01) |
| Mexican American | -0.01 | (0.01) |
| Puerto Rican American | -0.02* | (0.01) |
| Reading Difficulty Level | | |
| Intercept (White) | 0.13*** | (0.02) |
| African American | -0.08** | (0.03) |
| Mexican American | -0.10*** | (0.03) |
| Puerto Rican American | -0.04 | (0.03) |
| Response Format (Ref= Numeric) | | |
| Yes/No | | |
| Intercept (White) | -1.67*** | (0.20) |
| African American | 0.80** | (0.27) |
| Mexican American | 0.33 | (0.27) |
| Puerto Rican American | 0.30 | (0.29) |
| Vague Quantifier | | |
| Intercept (White) | -1.81*** | (0.20) |
| African American | 0.97** | (0.27) |
| Mexican American | 0.60* | (0.25) |
| Puerto Rican American | 0.72** | (0.26) |

*p<.05; **p<.01; *** p<.001.

Question length was found to have a positive main effect on comprehension difficulty: as the number of words increased, so did the likelihood that respondents would express comprehension problems. The negative coefficient associated with Puerto Rican status (-.02) in Table 4 indicated that difficulties with question comprehension increased with question length among non-Hispanic whites at a greater rate than among Puerto Ricans. Recognizing that Puerto Rican respondents in general expressed more comprehension difficulties than whites, it would appear that white respondents are more sensitive to variations in question length than are Puerto Ricans.

The reading difficulty level of survey questions was also found to be positively associated with comprehension difficulty: increased reading level was associated with more question comprehension difficulty. Race/ethnicity again moderated the effects of this variable. The negative regression coefficients associated with African-American and Mexican-American identities indicated that members of these two cultural groups experienced less additional comprehension difficulty as the reading level of survey questions increased, compared to White respondents. Although not significant, the regression coefficient associated with Puerto Rican ethnic identity suggests a similar relationship pattern. That is, reading level seems to be more problematic for White respondents (see Table 4).

Both main and interaction effects were also found for question response format. Specifically, question response formats that asked respondents to provide a numeric response (e.g. number of physician visits) generated more overall comprehension difficulties when compared with response formats in which respondents were asked to answer using a set of predefined response options (including both "yes-no" and vague quantifier formats). In addition, race/ethnicity was found to moderate the effects that the vague quantifier response format had on question comprehension. In particular, compared to Whites, African-American, Mexican-American, and Puerto Rican respondents were each more likely to express comprehension difficulties when vague quantifier response formats were employed. In addition, when 'yes-no' response formats were employed, African-American respondents were more likely to express comprehension difficulties, compared with Whites. The model in Table 4 was re-estimated to examine race/ethnic group differences in comprehension difficulties when asked questions requiring a numeric response (results not shown). No differences were found across groups in the likelihood of expressing comprehension difficulties with the numeric response format.

## 4 Discussion

This research used behavioral coding protocols for survey interviews to confirm cross-cultural differences in respondent's ability to comprehend a set of health-related survey items. Non-Hispanic Whites, in particular, expressed comprehension problems with a smaller number of survey questions, relative to the three minority groups also interviewed. These differences remained after controlling for other demographic characteristics, particularly age and education, that one might also expect to find associated with question comprehension. As the survey questions included were all selected from national health surveys in the United States, it is most likely the case that they were developed by representatives of the country's dominant non-Hispanic White culture. As such, it is not surprising that respondents from this group in general had less difficulty interpreting these questions.

Main effects of the four question characteristics evaluated were also identified. Questions deemed to be most abstract appear to be more likely to elicit comprehension difficulties among survey respondents, compared to those classified as least abstract. Likewise, comprehension problems also increase with the reading level and length of survey items. Response formats are additionally linked to comprehension problems, with questions that request numeric estimates generating more comprehension problem-related behaviors, compared to items that provide sets of response options. These findings are consistent with much of the available literature regarding question-based sources of processing error in survey research (Groves et al., 2004; Tourangeau, Rips & Rasinski, 2000). Findings related to race/ethnic differences in comprehension difficulties that are linked to specific question design features, however, have not been previously reported. Comprehension differences across groups were detected in responses to three of the four question design features examined: response format, question length, and reading level.

What is most perplexing in reviewing findings related to question response formats is that those formats that generate the fewest overall comprehension difficulties (i.e., the 'yes-no' and vague quantifier formats) ironically also generate the largest cross-cultural disparities. The greatest variability across groups was found in response to questions employing sets of vague quantifiers. The ambiguities associated with this response format are well known (Tourangeau, Rips & Rasinski, 2000). It is thus perhaps less surprising that cultural differences in comprehension difficulties were found among those questions employing sets of vague quantifiers as response options. We believe the same culture-based processes associated with main effect differences in comprehension difficulties across groups may also be responsible for these differences. Just as survey questions and instruments are largely prepared by White middle-class researchers in the U.S., so too are the response categories

attached to many of these items. It thus seems reasonable to assume that the processes responsible for general race/ethnic comprehension disparities in these data are also highlighted among those questions employing sets of vague quantifier response options, which are almost by definition subject to a greater range of interpretations than simple 'yes-no' formats.

Interestingly, one of the few available recommendations in regards to constructing cross-culturally equivalent survey items is the advice to use dichotomous response options ('yes-no,' 'agree-disagree') whenever possible (Smith, 2004). Our findings, however, indicate that some cultural variations in comprehension difficulties are also found among health questions employing binary 'yes-no' response options. Interestingly, no cross-group differences were found in difficulty answering numeric response format questions. Although this format elicits the greatest overall level of comprehension problems, it appears to produce little cross-group variability. That is, all respondents find numeric response formats equally and most difficult to answer. Of course, for many epidemiologic applications, numeric question response formats are unavoidable. Investigators should nonetheless exercise caution when requesting numeric information from survey respondents.

We also found differential effects of question reading level on respondent comprehension across race/ethnic groups. Specifically, increasing reading level was associated with more comprehension difficulty among White respondents, compared to African Americans and Mexican Americans. When examined in conjunction with the main effects of race/ethnicity on question comprehension, these findings suggest that questions with higher levels of reading difficulty may actually serve to *decrease* the gap in question comprehension between White and minority respondents. Overall, Whites are less likely to express comprehension problems. When confronted with questions at higher reading levels, however, Whites are more likely to exhibit comprehension problem behaviors, bringing them closer to the levels experienced by minority respondents in response to questions at all levels of reading difficulty. It would again appear that those questions that eliminate cross-cultural variability in comprehension (i.e., those worded at a high reading level) are also those that are most difficult for all respondents to comprehend. A similar pattern was found in regards to the cross-group effects of question length. Whereas shorter questions produced fewer comprehension difficulties, race/ethnic differences were minimized for longer questions.

Overall, these findings suggest that well-founded efforts to improve the general comprehension of survey questions may in some cases have the unintended effect of increasing cross-cultural variability. Simplifying question length, reading level and response formats would appear to improve overall question comprehension at the cost of enhancing cross-cultural disparities. These findings are perhaps a legacy of the process by which knowledge of survey question design features has accumulated over the past 50 years in the U.S.

This research has largely over-represented White, non-Hispanic survey respondents (Payne, 1951; Schuman & Presser, 1981; Sudman & Bradburn, 1982) and it should perhaps thus not be surprising that findings are less generalizable to respondents from other cultural backgrounds. Clearly, more work is essential to identify question design features that effectively minimize comprehension problems in general and which also eliminate disparities in comprehension across cultural groups.

We wish to qualify our findings by acknowledging several important limitations. These include the fact that only a relatively small sample of health-related survey questions (n=45) were examined. Future work will need to evaluate larger and more diverse sets of survey questions. Second, behavioral coding has not been previously used for the purposes described in this paper. This methodology was initially developed as a method of evaluating interviewer performance (Cannell et al., 1975) and has been previously used both to evaluate the performance of survey questions (Oskenberg et al., 1991; van der Zouwen & Smit, 2004) and survey interviewers (Dykema, Lepkowski & Blixt, 1997). We believe, however, that this approach to assessing cultural variations in question understanding has good face validity and is an appropriate use of these methods. Our sample of respondents was relatively small and non-random, raising questions about the generality of our findings. Finally, we also note that findings should not be generalized beyond the U.S. context. Nonetheless, this work is based on more than 13,000 survey responses by members of four distinct race/ethnic groups. In addition, the findings are consistent with a growing body of research literature demonstrating the existence of cross-cultural variations in the comprehension and interpretation of survey questions. The development of methods and procedures for establishing the conceptual equivalence of survey measures across cultures should be encouraged as the U.S. continues to evolve into a more culturally heterogeneous society.

## References

Andrews, F. M. 1984. "Construct validity and error components of survey measures: A structural modeling approach." *Public Opinion Quarterly* 48:409-442.

Bradburn, N. M., and C. Miles. 1979. "Vague quantifiers." *Public Opinion Quarterly* 43:9-101.

Bradburn, N. M., and S. Sudman. 1979. *Improving Interview Method and Questionnaire Design.* San Francisco: Jossey-Bass.

Brislin, R. W. 1986. "The wording and translation of research instruments." Pp. 137-164 in *Field Methods in Cross-Cultural Research,* edited by Lonner, W. J., and J. W. Berry. Beverly Hills: Sage.

Cannell, C. F., S. A. Lawson, and D. L. Hausser. 1975. *A Technique for Evaluating Interviewer Performance.* Ann Arbor: University of Michigan.

D'Andrade, R., N. Quinn, S. Nerlove, and A. K. Romney. 1972. "Categories of disease in American-English and Mexican-Spanish." Pp. 9-54 in *Multidimensional Scaling,* volume II, edited by Romney, A. K., R. Shepard, and S. B. Nerlove. New York: Seminar Press.

Dykema, J., J. M. Lepkowski, and S. Blixt. 1997. "The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study." Pp. 287-310 in *Survey Measurement and Process Quality,* edited by Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, et al. New York: John Wiley & Sons.

Flesch, R. 1979. *How to Write Plain English.* New York: Harper and Row.

Fowler, F. J. 1995. *Improving Survey Questions: Design and Evaluation.* Thousand Oaks: Sage.

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology.* New York: John Wiley & Sons.

Harkness, J. A. 2003. "Questionnaire translation." Pp. 35-56 in *Cross-Cultural Survey Methods,* edited by Harkness, J. A., F. J. R. van de Vijver, and P. Ph. Mohler. New York: John Wiley & Sons.

Hedeker, D., and R. Gibbons. 1994. "A random-effects ordinal regression model for multilevel analysis." *Biometrics* 50:993-944.

Jenkins, J. H. 1988. "Conceptions of schizophrenia as a problem of nerves: A cross-cultural comparison of Mexican-Americans and Anglo-Americans." *Social Science and Medicine* 26:1233-1243.

Johnson, T. P., D. O'Rourke, N. Chavez, S. Sudman, R. Warnecke, L. Lacey, et al. 1997. "Social cognition and responses to survey questions among culturally diverse populations." Pp. 87-113 in *Survey Measurement and Process Quality,* edited by Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, et al. New York: John Wiley & Sons.

Johnson, T. P., D. O'Rourke, N. Chavez, S. Sudman, R. B. Warnecke, L. Lacey, et al. 1996. "Cultural variations in the interpretation of health survey questions." Pp. 57-62 in *Health Survey Research Methods Conference Proceedings,* edited by Warnecke, R. B. DHHS Publication No. (PHS) 96-1013. Hyattsville, MD: National Center for Health Statistics.

Just, M. A., and Carpenter, P. A. 1992. "A capacity theory of comprehension: Individual differences in working memory." *Psychological Review* 99:122-149.

Knäuper, B., R. F. Belli, D. H. Hill, and A. R. Herzog. 1997. "Question difficulty and respondents' cognitive ability: The effect on data quality." *Journal of Official Statistics* 13:181-199.

Laurent, A. 1972. "Effects of question length on reporting behavior in the survey interview." *Journal of the American Statistical Association* 67:298-305.

McKay, R. B., M. J. Breslow, R. L. Sangster, S. M. Gabbard, R. W. Reynolds, J. M. Nakamoto, et al. 1996. "Translating survey questionnaires: Lessons learned." *New Directions for Evaluation* 70:93-104.

Meredith, L. S., and A. L. Siu. 1995. "Variation and quality of self-report health data: Asians and Pacific Islanders compared with other ethnic groups." *Medical Care* 33:1120-1131.

Morse, J. M., and R. M. Morse. 1988. "Cultural variation in the inference of pain." *Journal of Cross-Cultural Psychology* 19:232-242.

Oksenberg, L., C. Cannell, and G. Kalton. 1991. "New strategies for pretesting survey questions." *Journal of Official Statistics* 7:349-365.

Payne, S. L. 1951. *The Art of Asking Questions.* Princeton: Princeton University Press.

Raudenbush, S. W., A. S. Bryk, Y. Fai Cheong, and R. T. Congdon. 2004. *HLM 6: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood, IL: Scientific Software International, Inc.

Schuman, H., and S. Presser. 1981. Questions and Answers in Attitude Surveys. San Diego: Academic Press.

Smith, T. W. 2004. "Developing and evaluating cross-national survey instruments." Pp. 431-452 in *Methods for Testing and Evaluating Survey Questionnaires,* edited by Presser, S., J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin et al. Hoboken: John Wiley.

Sperber, A. D., R. F. DeVellis, and B. Boehlecke. 1994. "Cross-cultural translation: Methodology and validation." *Journal of Cross-Cultural Psychology* 25:501-524.

Sudman, S., and N. Bradburn. 1982. *Asking Questions.* San Francisco: Jossey-Bass.

Teresi, J., D. Holmes, M. Ramierz, B. Gurland, and R. Lantigua. 2001. "Performance of cognitive tests among different racial/ethnic and education groups: Findings of differential item functioning and possible item bias." *Journal of Mental Health and Aging* 7:79-89.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

van der Zouwen, J., and J. H. Smit. 2004. "Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach." Pp. 109-130 in *Methods for Testing and Evaluating Survey Questionnaires,* edited by Presser, S., J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, et al. Hoboken: John Wiley.

Warnecke, R. B., T. P. Johnson, N. Chavez, S. Sudman, D. P. O'Rourke, L. Lacey, et al. 1997. "Improving question wording in surveys of culturally diverse populations." *Annals of Epidemiology* 7:334-342.